

2.1

VALIDITY AND RELIABILITY: DEFINING THE PROBLEM

This session is entitled 'Validity and reliability'. In New Orleans and Montreal we were told two things about these twin subjects.

One We were told what they mean

Two We were told that we did not have any!

In this brief introduction I quickly review the concepts of validity and reliability and how they relate to the issues we are addressing.

First and easier is reliability. Reliability means consistency: not accuracy, but consistency. Consistency has been measured in many ways. We have looked at the consistency of audience estimates across rotations, contractors, variations in both question and answer categories etc. Most often we have found a lack of consistency. But not always.

Validity is more complicated. First, it requires consistency. But it requires more. It requires accuracy. For a measure to be valid it must provide results which correspond to the real world. Validation studies require:

That we determine by observation or by some other empirical method that reading has occurred.

This known readership is then contrasted with readership claims.

Unfortunately, there have been few such studies.

Politz made the distinction between reliability and validity by saying an elevator can be called a reliable transportation instrument if, when you press a button, it always stops within, say, plus or minus one-fourth of an inch of floor level. However, it is not a valid transportation device if you never know what floor it will stop at.

Confusion between validity and reliability seemed to exist in Montreal in two areas:

(1) The assumption was sometimes made, that if between-publication bias is removed then our measures are valid, when in fact we may still have estimates which are far from actual readership levels.

(2) Assumptions about the validity of a method were made, based on how well its estimates agree with estimates of other unvalidated measures. For example, recency readership levels are evaluated based on their consistency with frequency claims or readership levels are assumed to be inaccurate because circulation estimates from the same survey are inaccurate, etc.

In these and other instances the concept of validity is side-stepped as one unknown is substituted for another unknown.

The Americans and the British as well as others are often at opposite poles when the question is, "Should we worry about validity? Or should we worry about reliability?"

A typical British position on validity is that it *cannot* be measured while a typical American position is that it *must* be measured.

Reasonably prevalent among British researchers is the notion that actual readership levels can never be known, and that even close approximations are not possible. Therefore they feel we should restrict our efforts to improving reliability.

Reasonably prevalent among American researchers is the idea that we must attempt to solve the validity problem. This is a traditional American position on all media research. Golden yardsticks such as the coincidental method have been used to validate television and radio audience

measurements and the glue-spot technique used to validate reports on magazine page openings.

Clearly it is most desirable to define and measure readership with a method which, with reasonable accuracy, yields audience estimates that fall in the same ballpark as 'truth'. We may ultimately decide that truth is beyond our grasp. We may conclude that readers are so diverse and the reading event is so ephemeral that we cannot come suitably close to estimating average issue audience levels. We may have to settle on selecting a method which yields consistent rather than accurate results. We would then proceed to improve our methods to eliminate or resolve all known

inconsistencies and biases.

Another approach would be to replace average issue audience with some other measure such as reading days or magazine page exposure. Presumably we can measure these dimensions more accurately than average issue audience.

But it is premature to give up on the concepts of audience reach and frequency and settle only for a measure of gross rating points. In fact, making any of these compromises at this time is, in my judgement, extremely premature. So I am pleased that here we do not have to settle for a session that only covers reliability. We have one that covers validity as well.